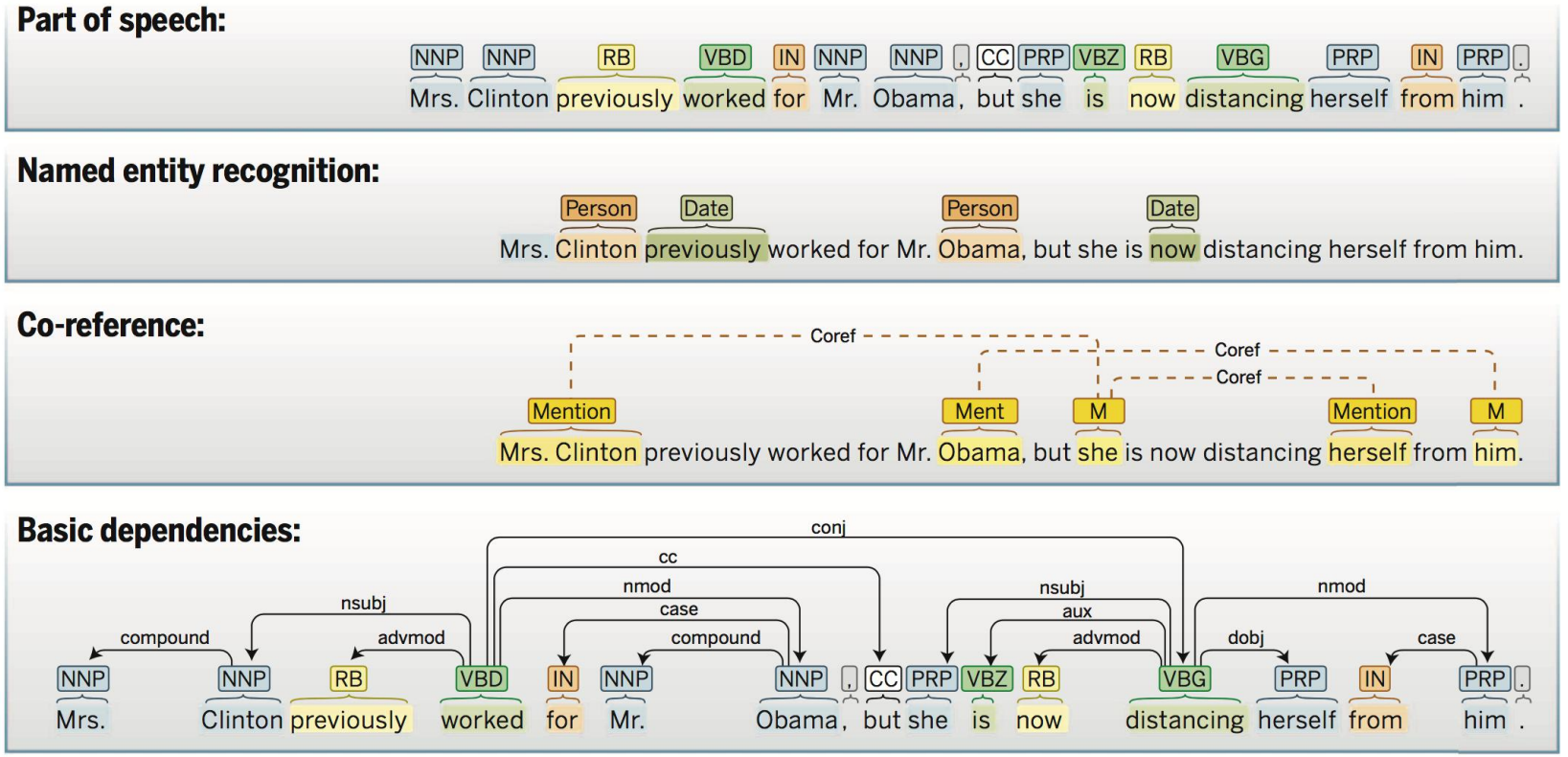
Knowledgeable Pretrained Language Mode

刘 知 远
清华大学

2020年12月19日

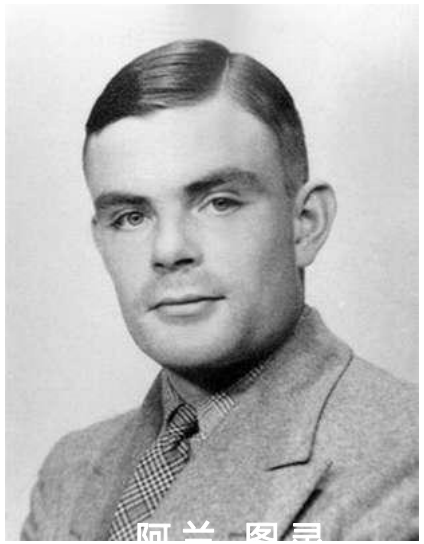
Natural Language Processing

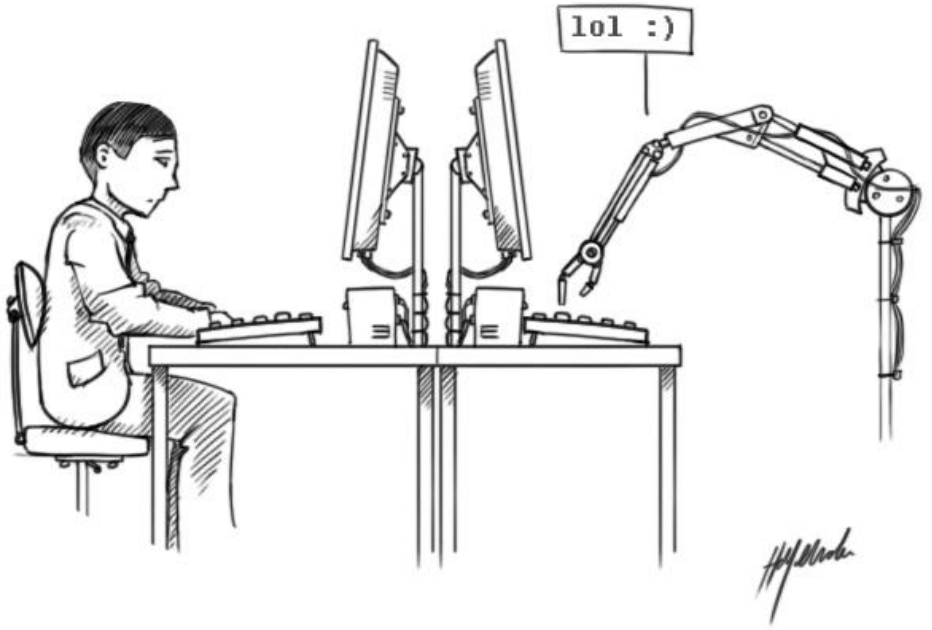
- NLP aims to make computers understand languages
- The nature of NLP is structure prediction

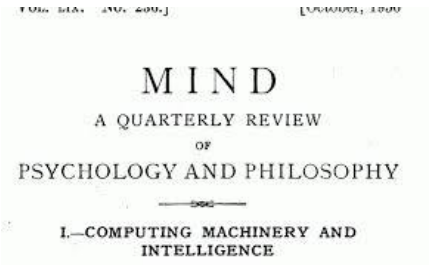


Advances in Natural Language Processing. Science 2015.

NLP Is The Key of AI






MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

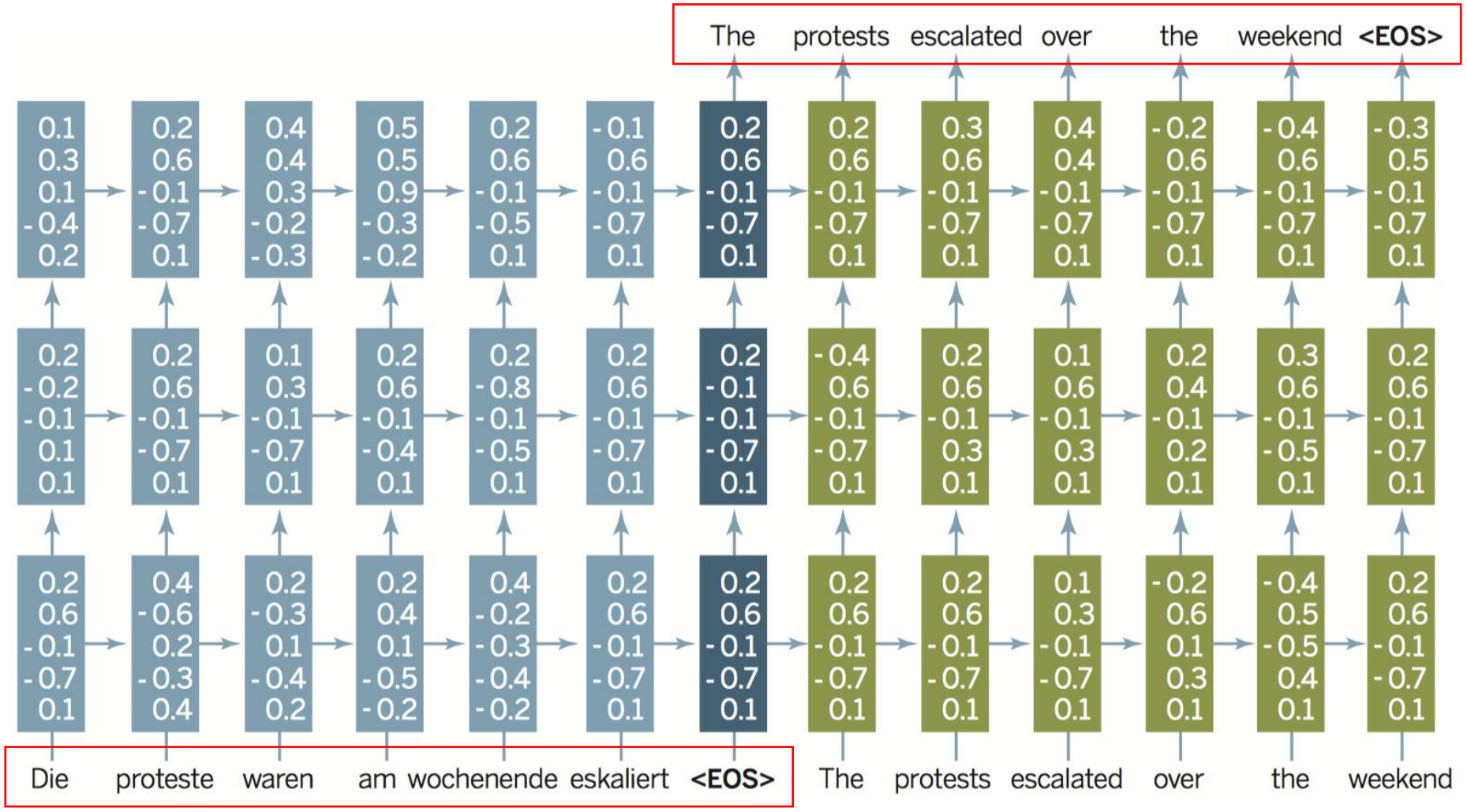

汉译世界学术名著丛书

计算机与人脑

〔美〕约·冯·诺意曼 著

NLP：The Key to Pass Turing Test and Realize AI

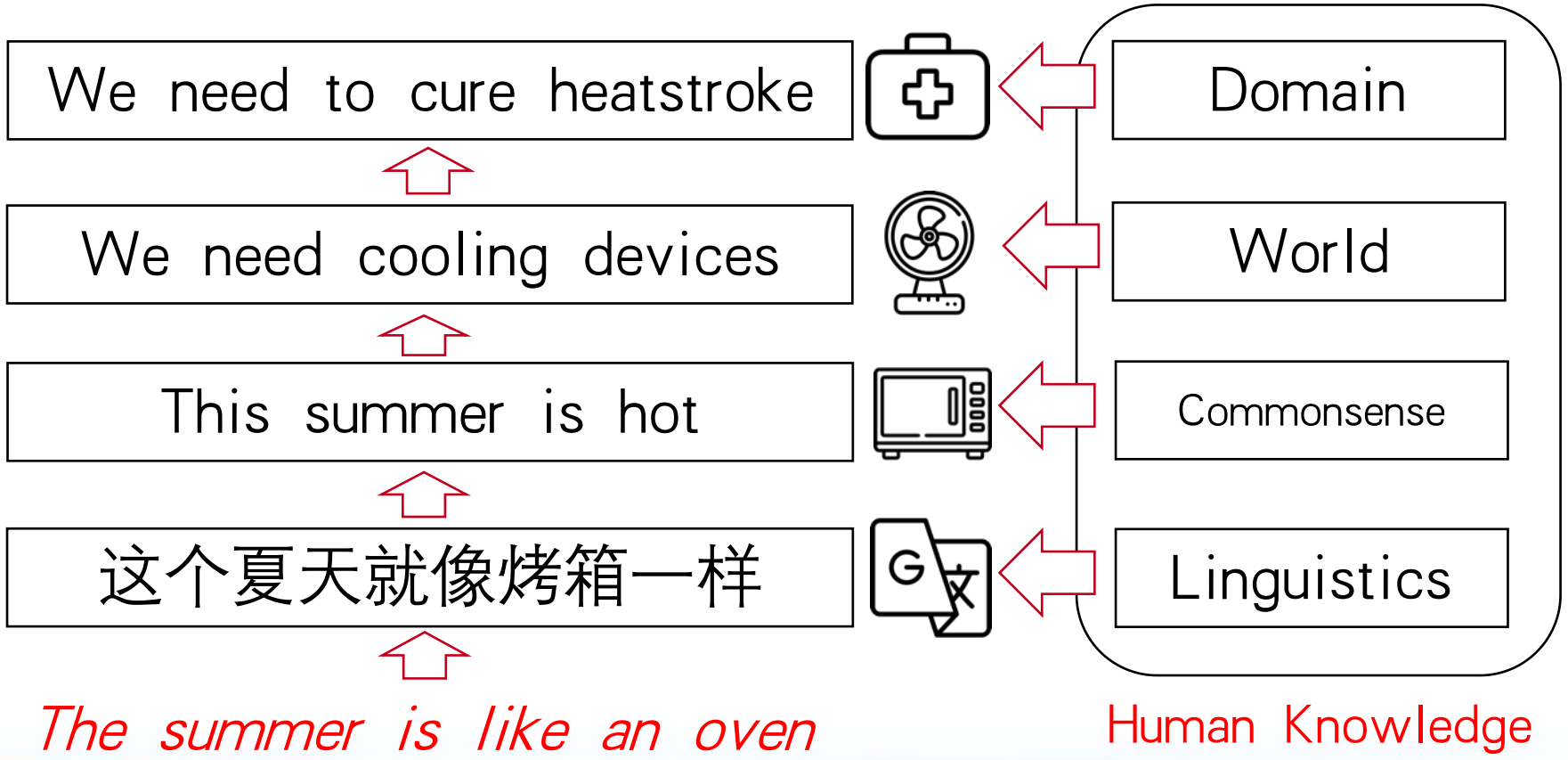
Deep Learning: Data-Driven NLP



Advances in Natural Language Processing. Science 2015.

Language & Knowledge

- Knowledge enables people to understand language from superficial meanings to implicative meanings



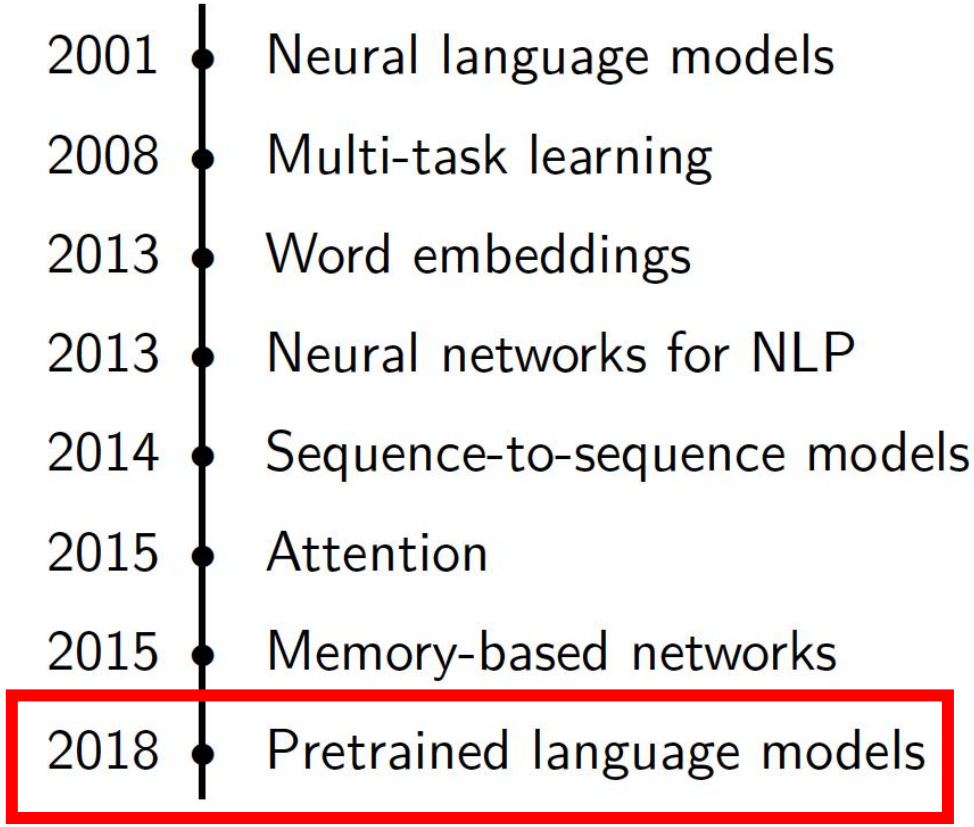
Challenges of DL for NLU & NLP

… we feel confident that more data and computation, in addition to recent advances in ML and deep learning, will lead to further substantial progress in NLP. However, the truly difficult problems of semantics, context, and knowledge will probably require new discoveries in linguistics and inference.

Advances in Natural Language Processing. Science 2015.

Pretrained Language Model as a Breakthrough in 2018

- Impressive progress of deep learning on unsupervised text corpora

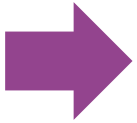


Sebastian Ruder http://ruder.io/a-review-of-the-recent-history-of-nlp/

What is Language Model

- Language models aims to predict the probability of a sequence as a natural sentence, or predict the probability of the next word given context
- Language models are a key to NLP and semantic representation of documents

她是中国人工智能领域的著名_____

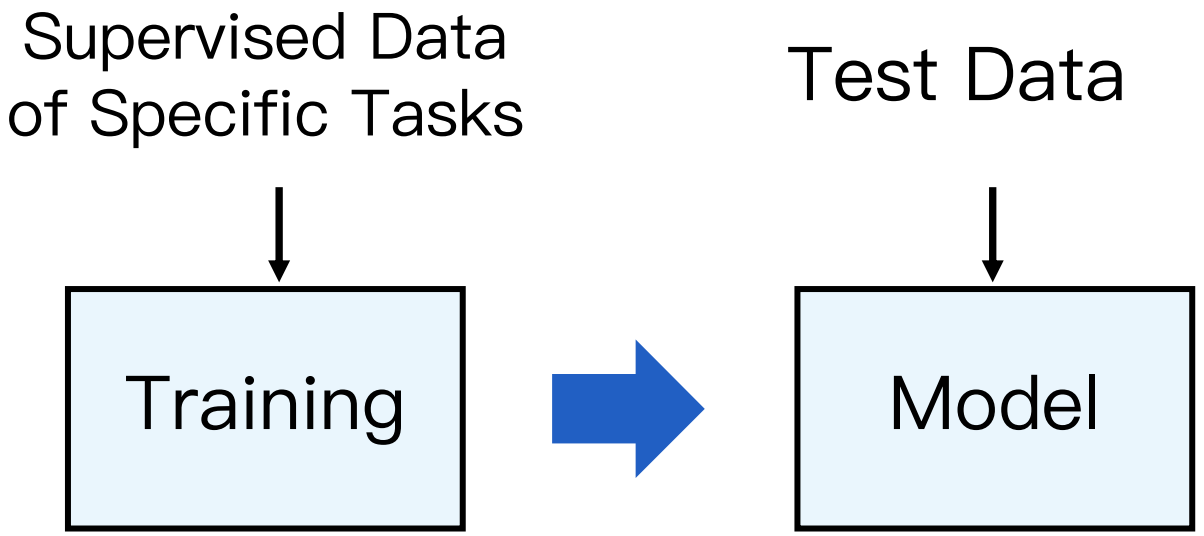


Language Model

Rank 1：专家
Rank 2：学者
Rank 3：科学家
Rank 4：教授

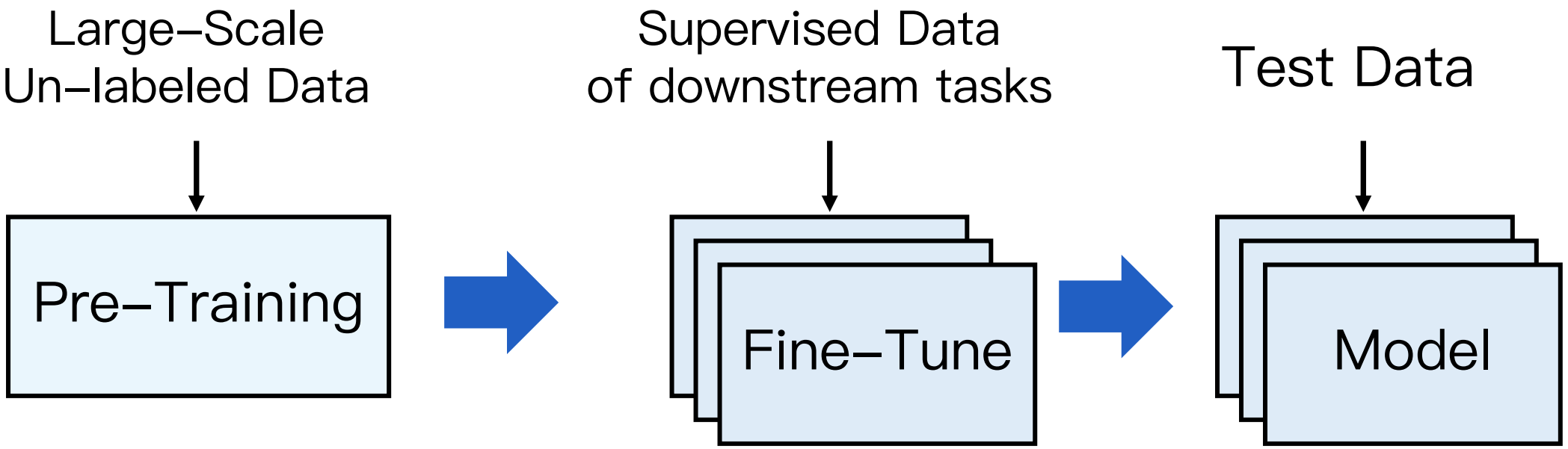
Challenge of Deep Learning in NLP

- Deep Learning has achieved the best performance in most NLP tasks
- Challenges: require large-scale supervised training

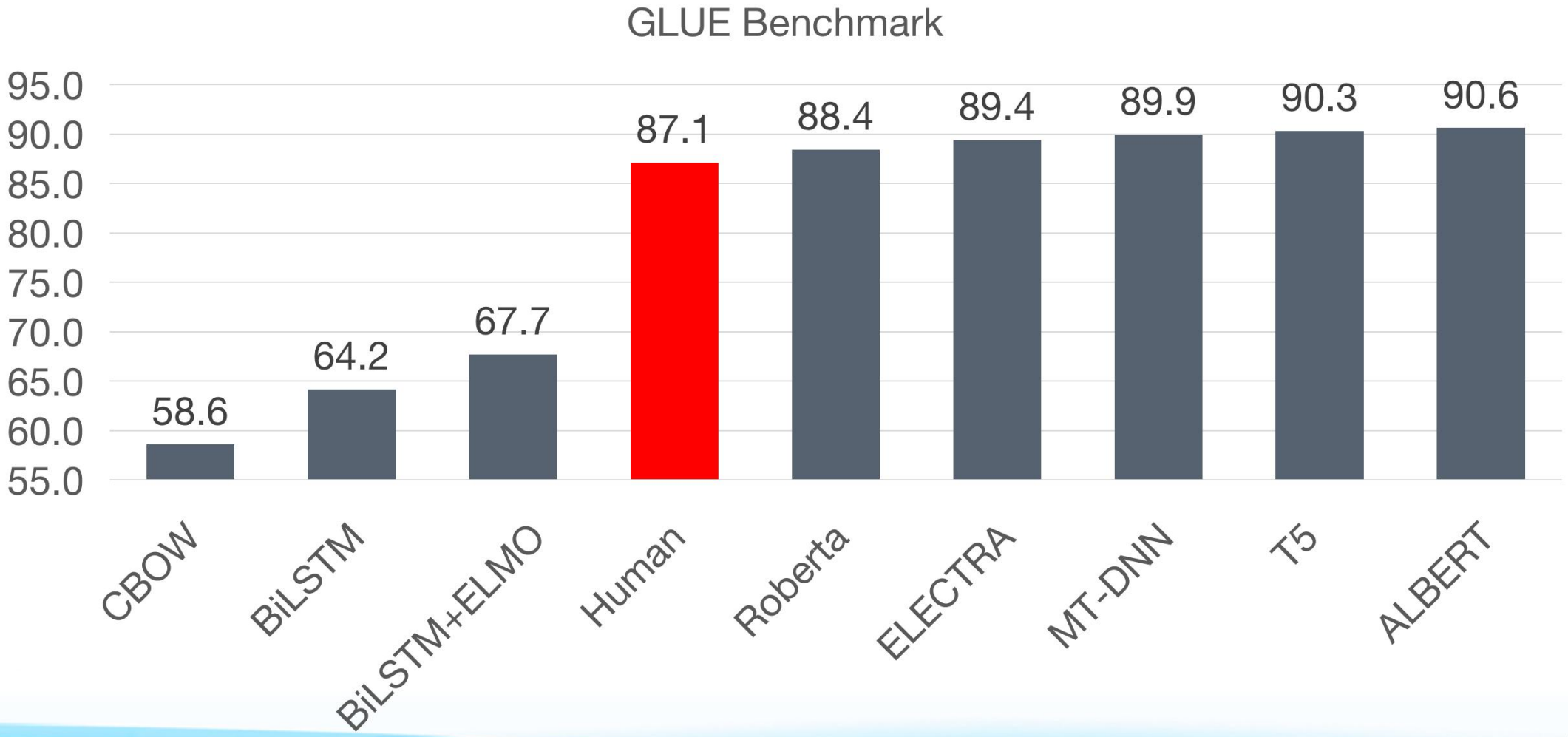


Pretrained Language Models

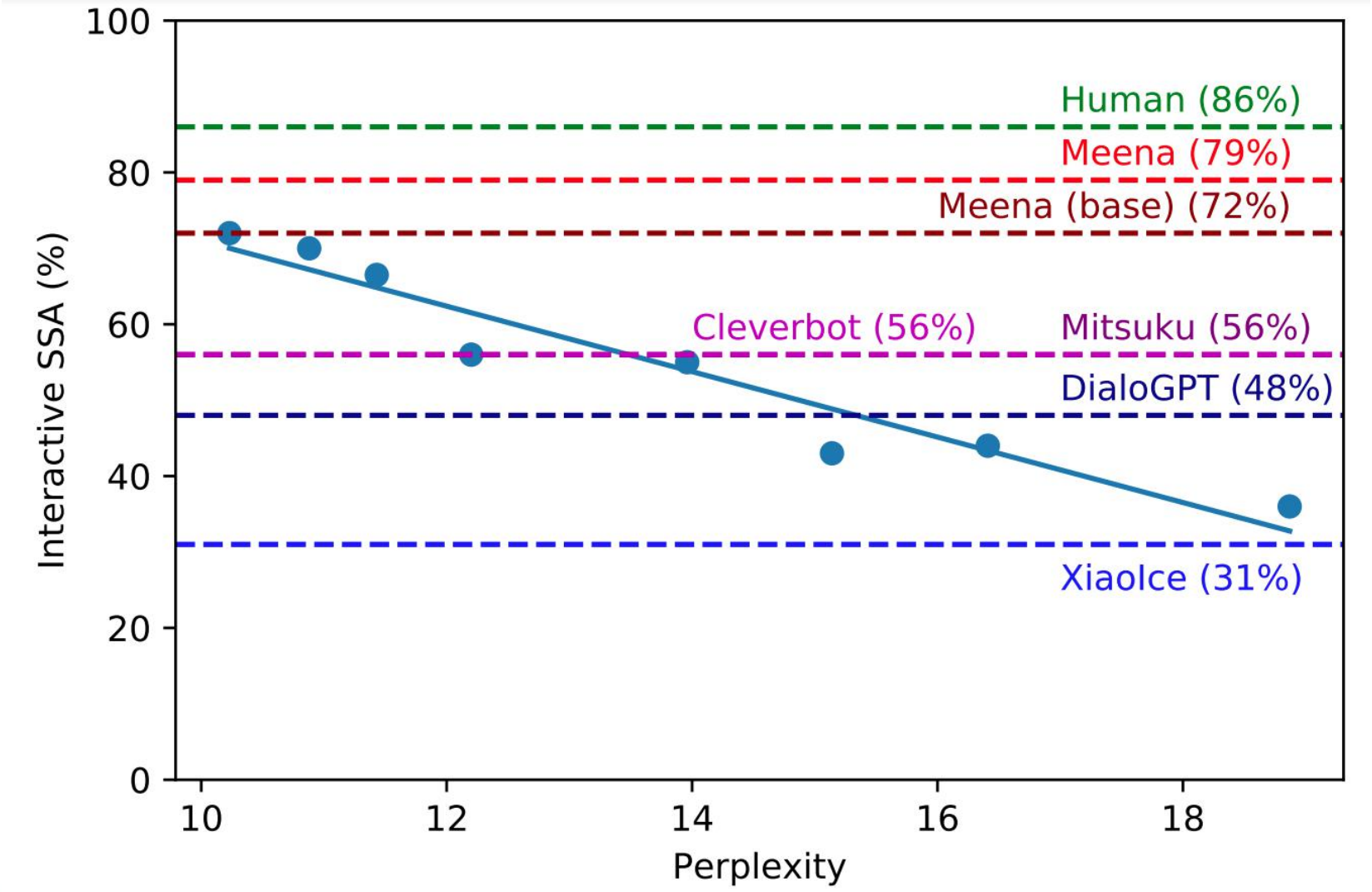
- Pre-trained Language Models (PLMs) can learn language patterns from large-scale un-labeled data, and improve the performance on downstream tasks by fine-tuning parameters



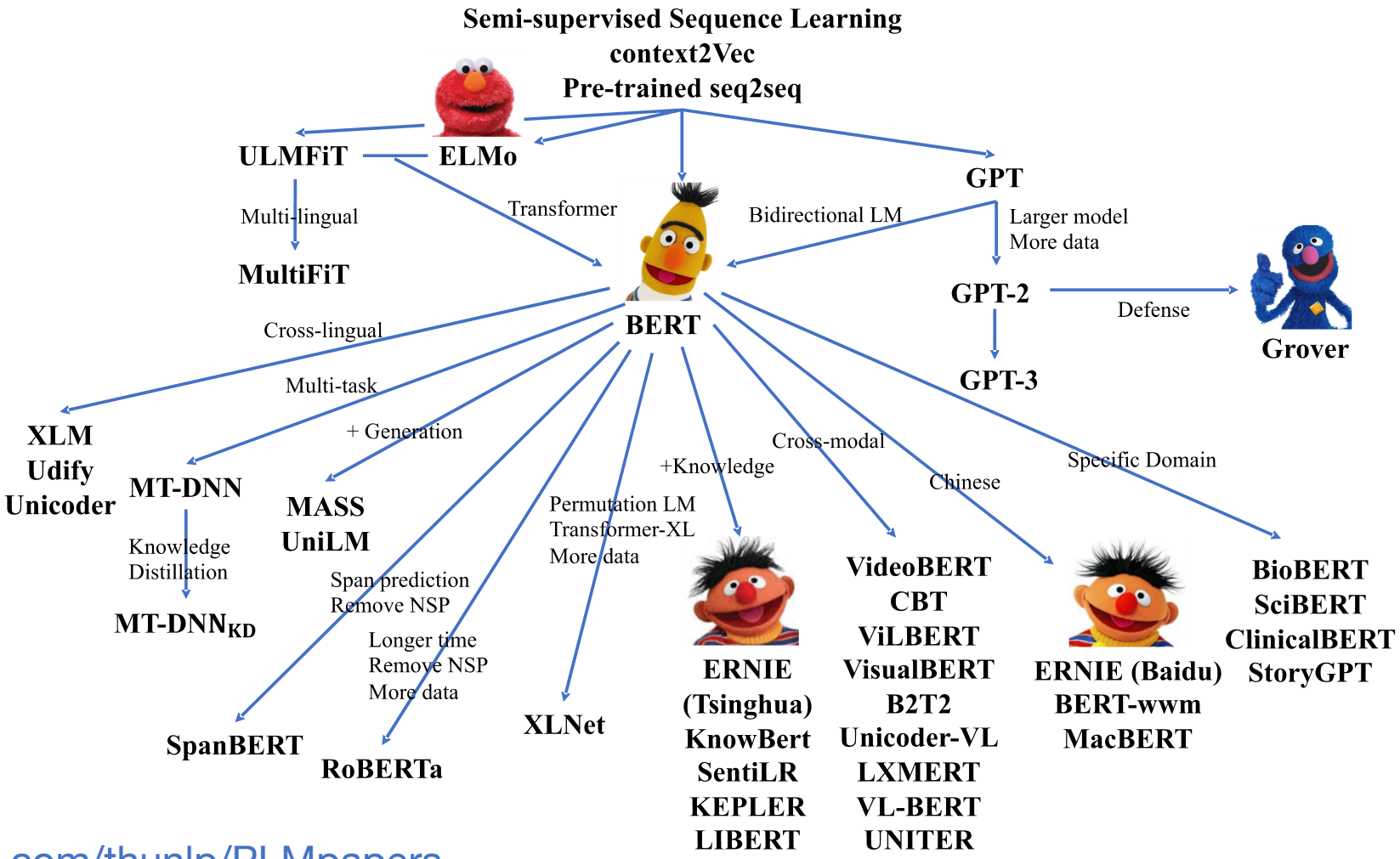
Superior Performance on Language Understanding



Superior Performance on Language Generation

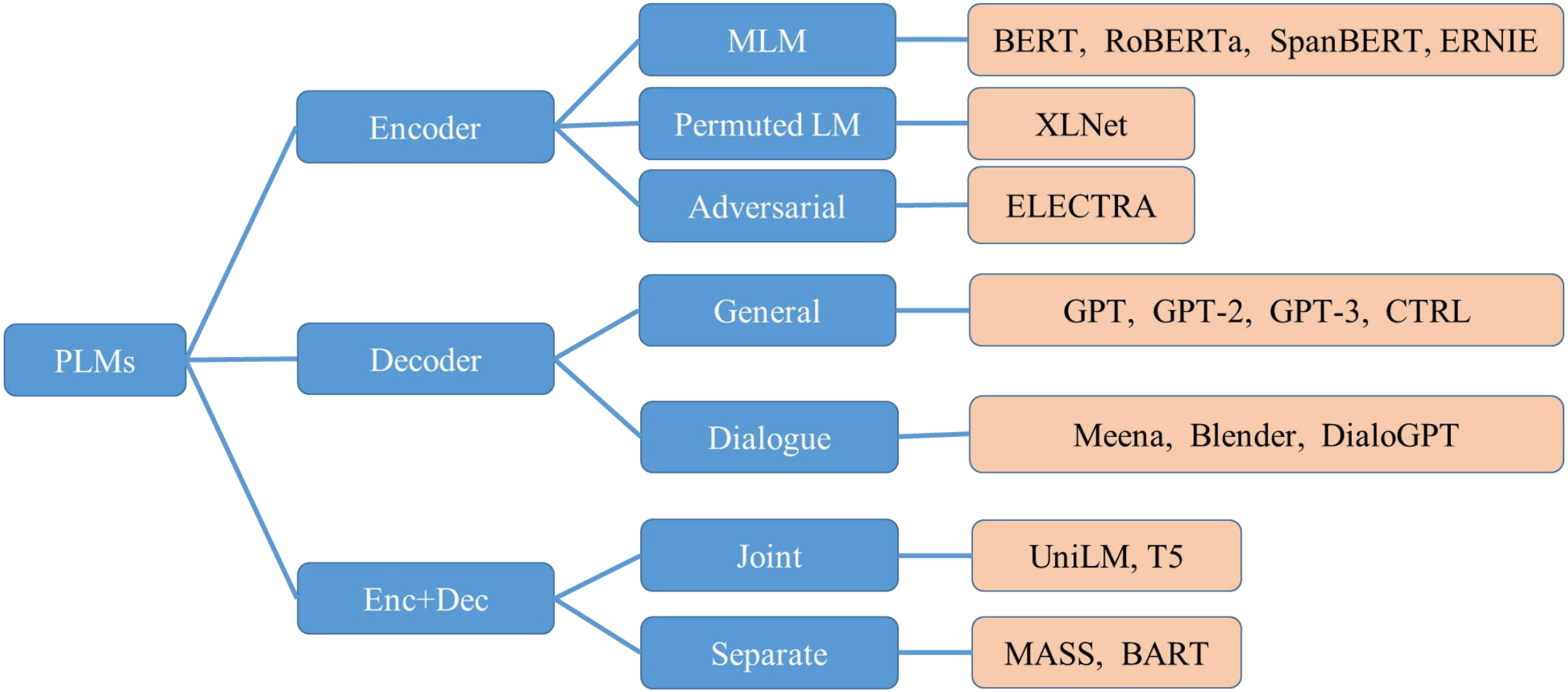


Contests of Pretrained Language Models



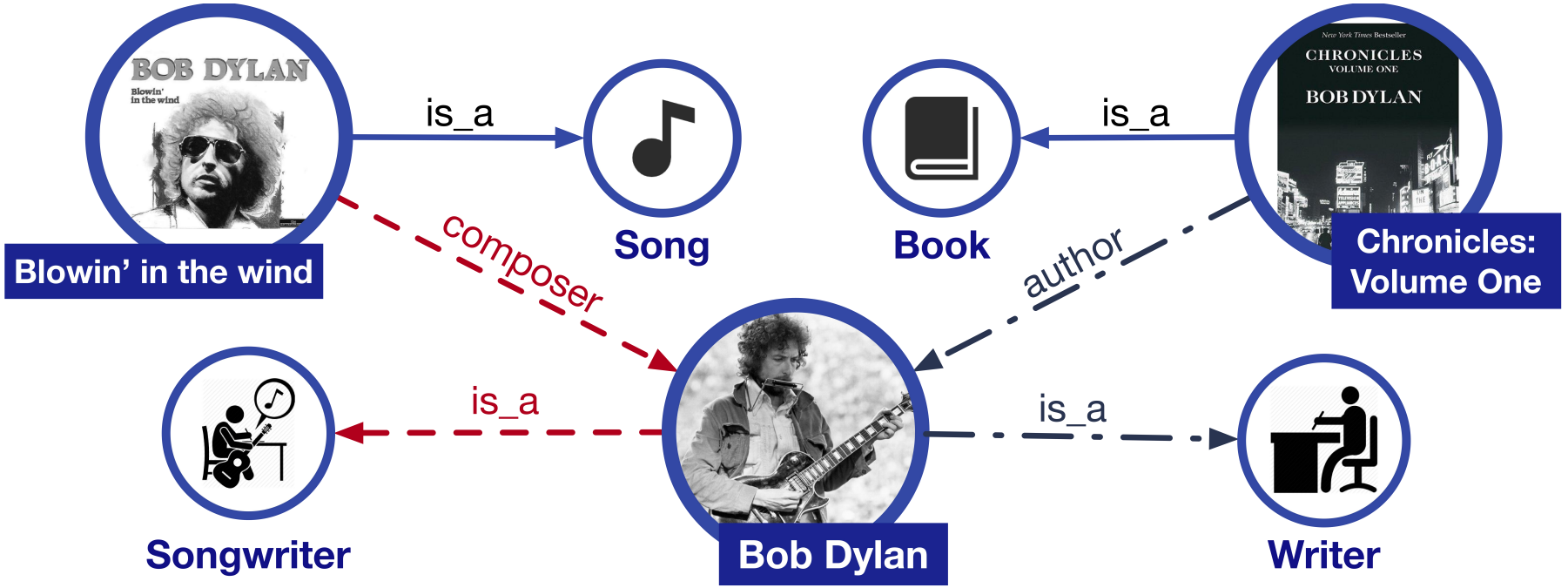
https://github.com/thunlp/PLMpapers

Contests of Pretrained Language Models



Knowledgeable PLM

- External knowledge information can benefit language understanding, for low resource entities, and implicit background knowledge



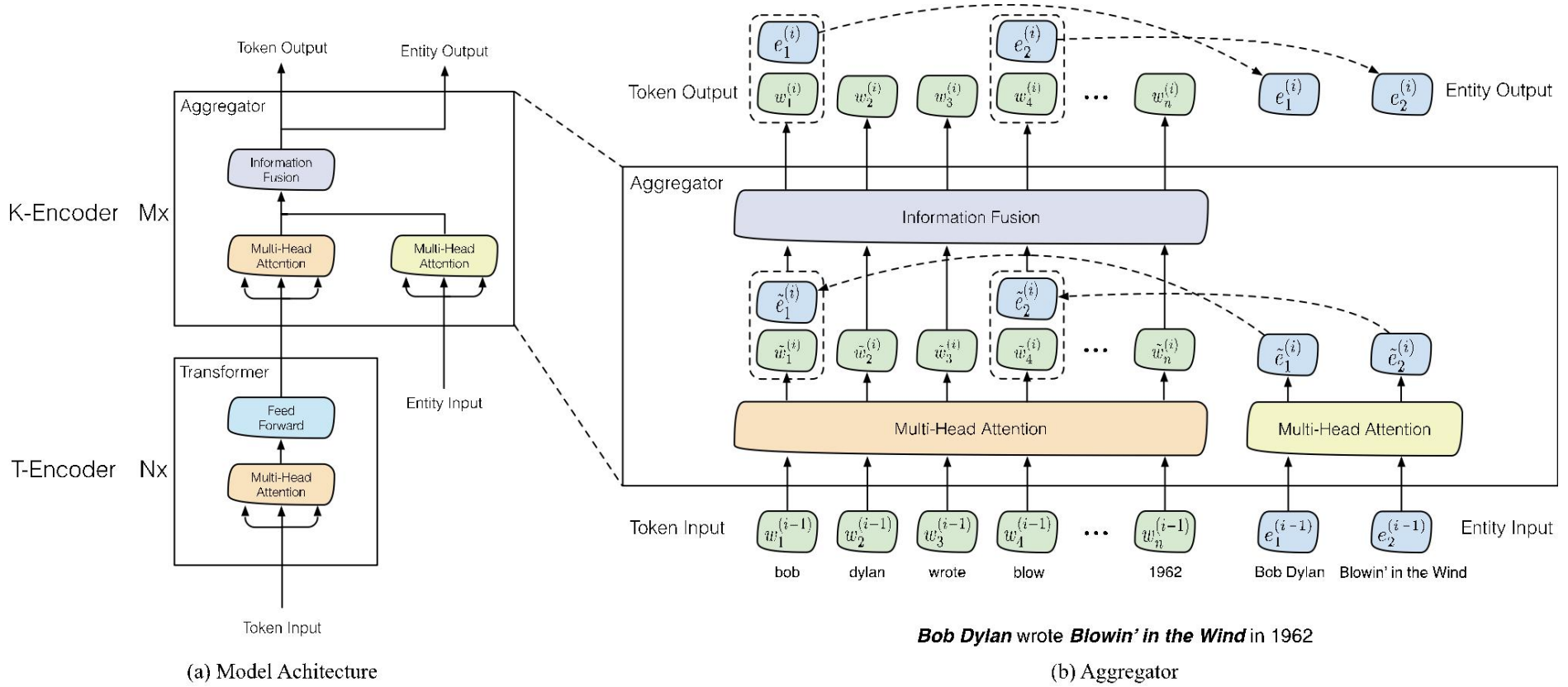
Bob Dylan wrote ***Blowin' in the Wind*** in 1962, and wrote ***Chronicles: Volume One*** in 2004.

How to Make PLMs Knowledgeable

- **Knowledgeable Input:** input augmentation as extra features
- **Knowledgeable Tasks:** knowledge-guided pre-training tasks
- **Knowledgeable Framework:** knowledge-guided neural architecture

Knowledgeable Input

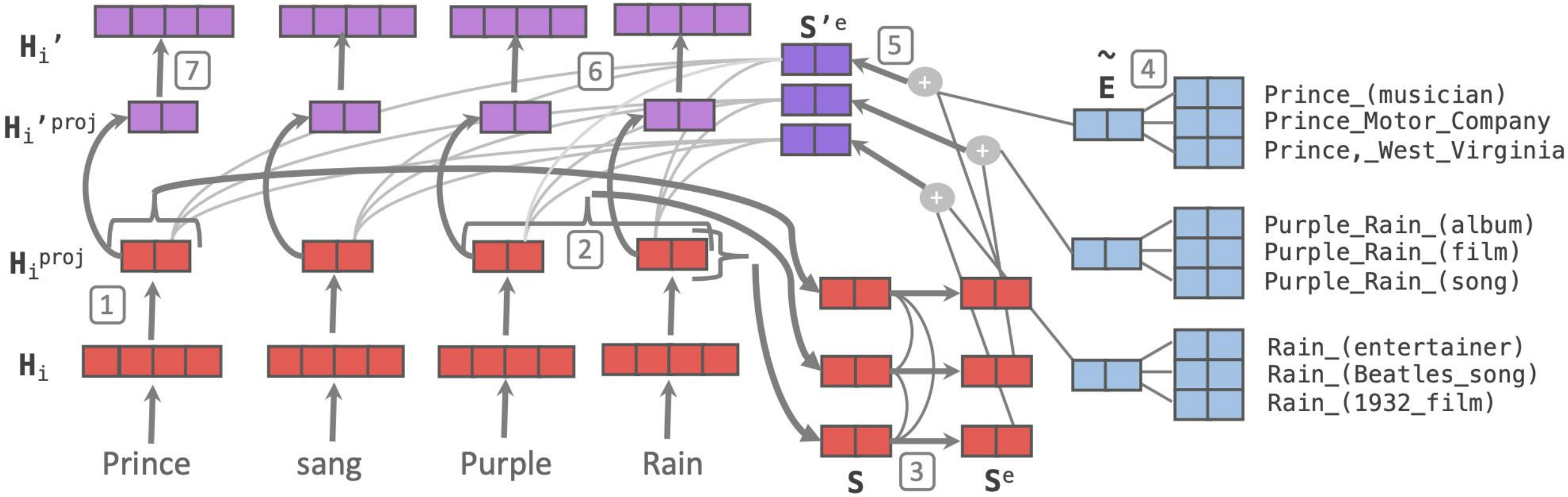
- ERNIE: Enhanced Language Representation with Informative Entities
  - Lower layers for text, and higher layers for knowledge integration



Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, Qun Liu. ERNIE: Enhanced Language Representation with Informative Entities. ACL 2019.

Knowledgeable Input

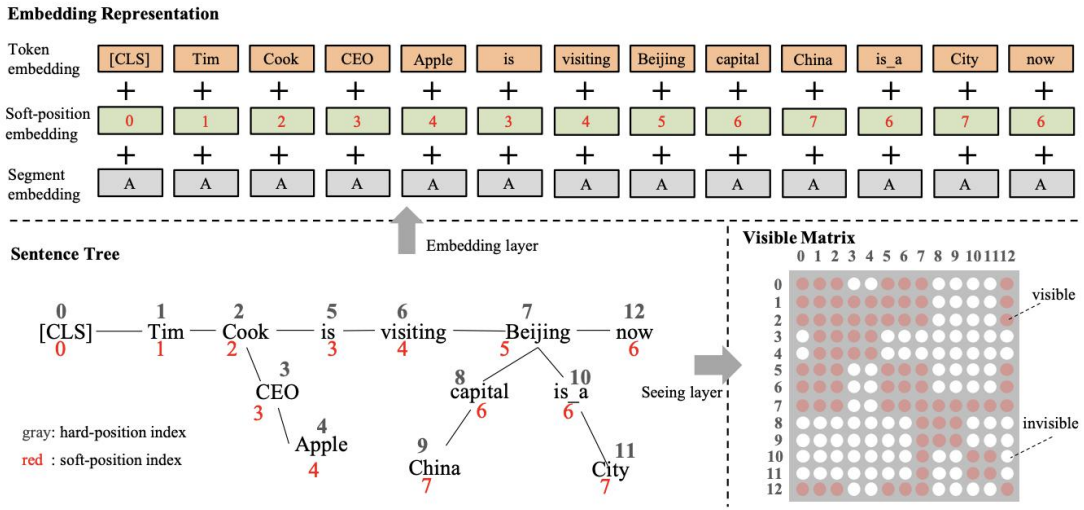
- KnowBERT: Knowledge Enhanced Contextual Word Representations

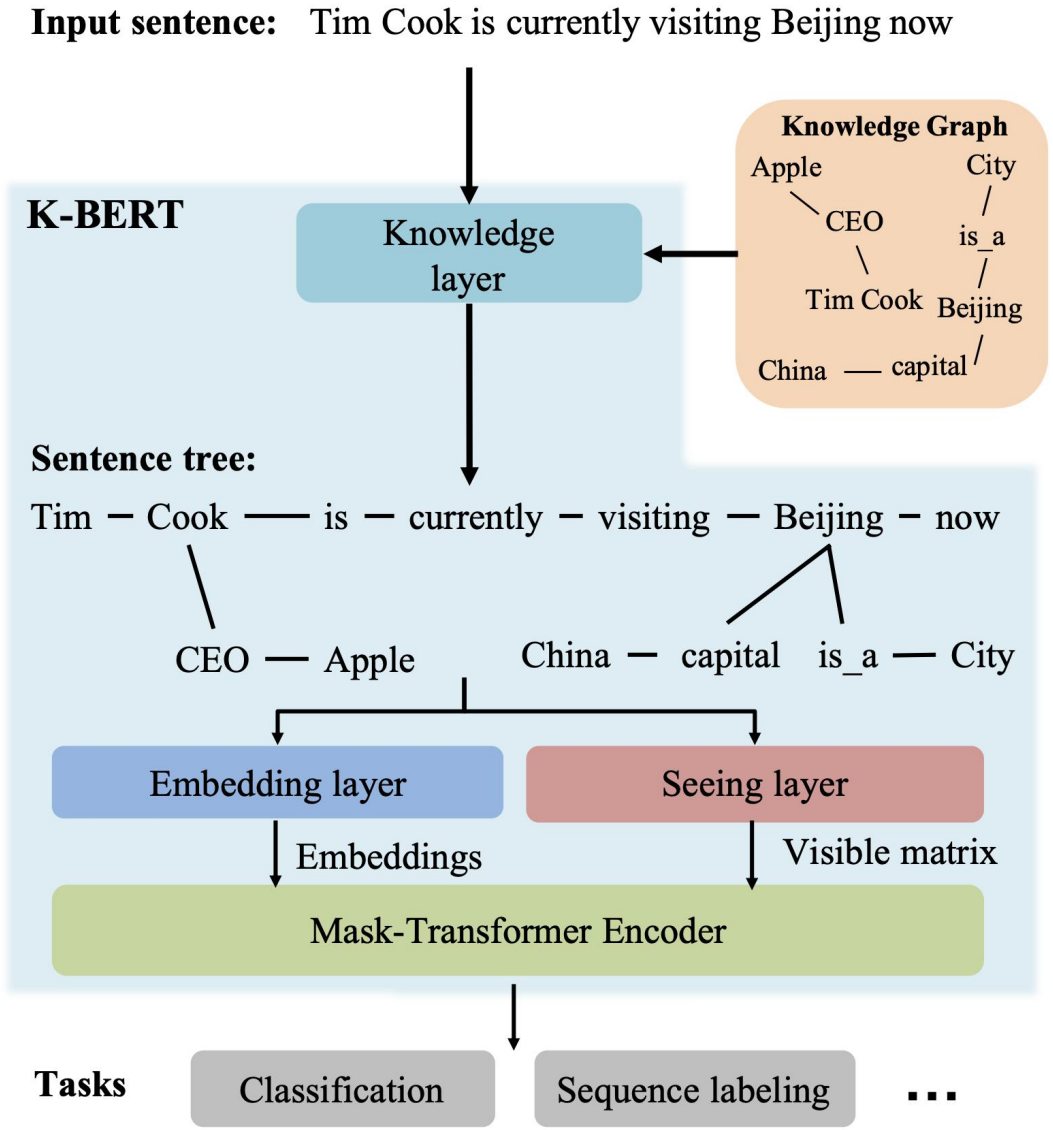


Matthew E. Peters, Mark Neumann, Robert Logan, Roy Schwartz, Vidur Joshi, Sameer Singh, Noah A. Smith. Knowledge enhanced contextual word representations. EMNLP-IJCNLP 2019.

Knowledgeable Input

- K-BERT: Directly add knowledge without further pre-training using knowledge layer

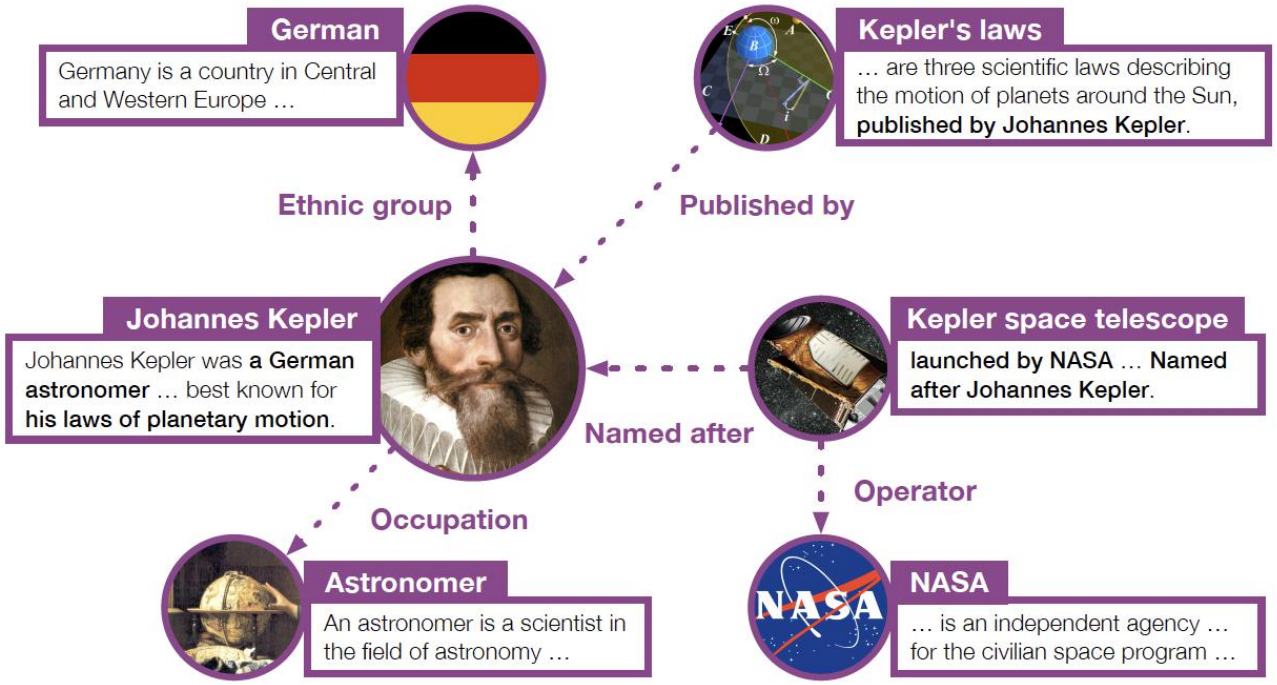




Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, Ping Wang. K-BERT: Enabling Language Representation with Knowledge Graph. AAAI 2020.

Knowledgeable Tasks

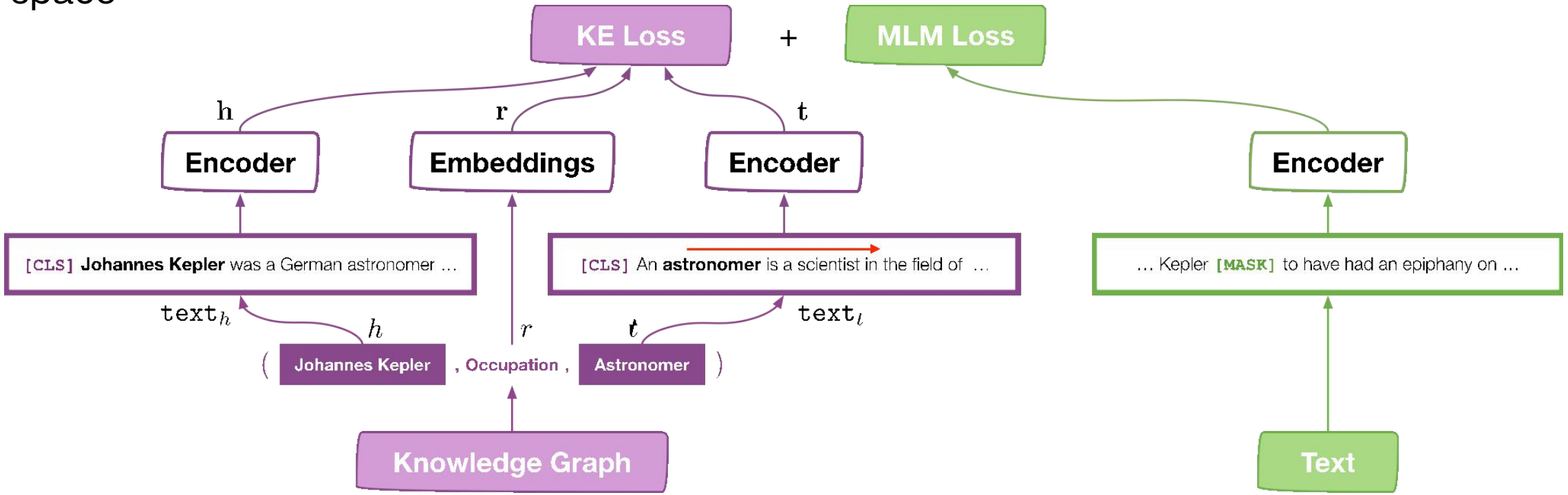
- KEPLER: Joint learning of knowledge and language modeling
- Unify knowledge embedding and language representation into the same semantic space



Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhengyan Zhang, Zhiyuan Liu, Juanzi Li, Jian Tang. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation. To appear at TACL.

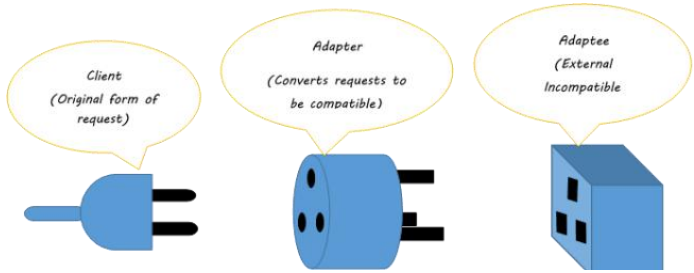
Knowledgeable Tasks

- KEPLER: Joint learning of knowledge and language modeling
- Unify knowledge embedding and language representation into the same semantic space

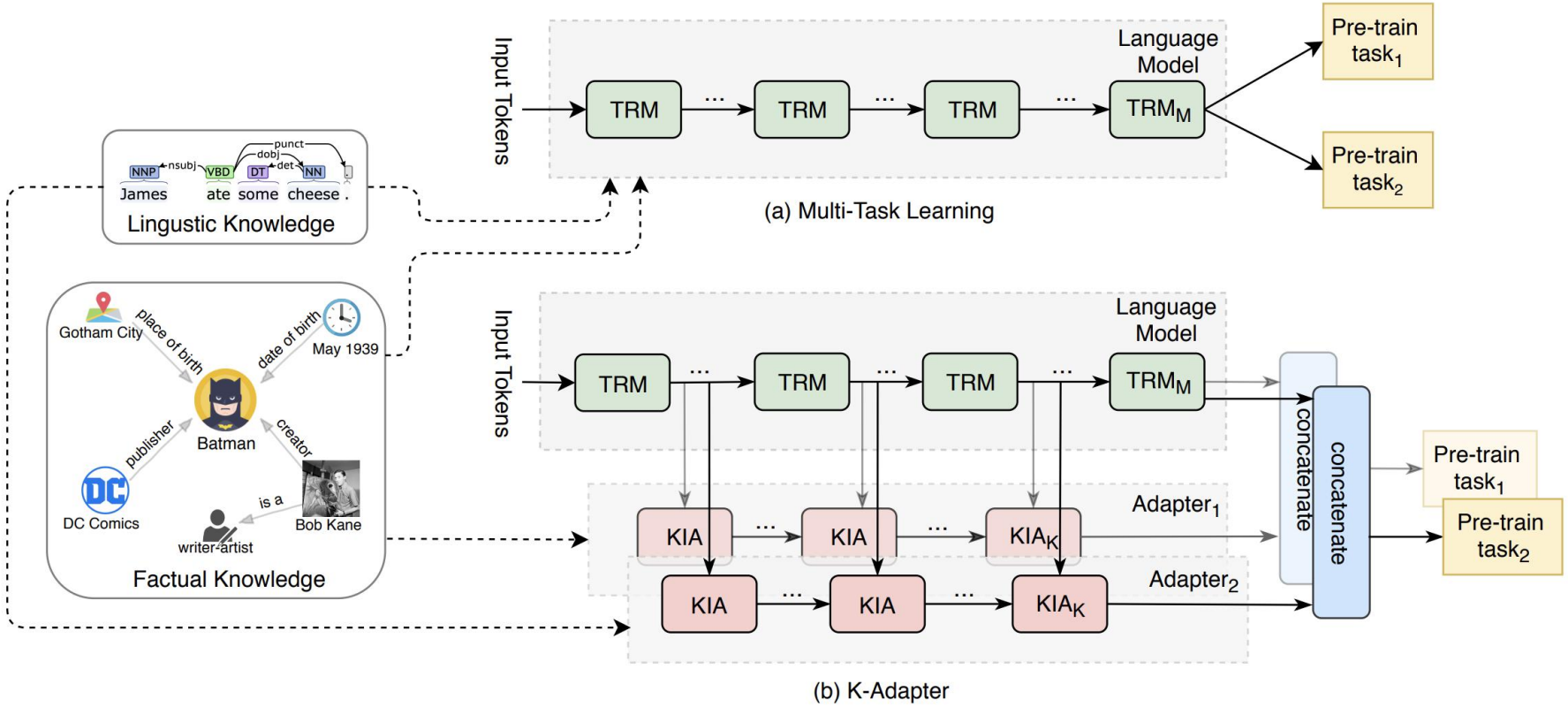


Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhengyan Zhang, Zhiyuan Liu, Juanzi Li, Jian Tang. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation. To appear at TACL.

Knowledgeable Framework



- K-Adapter: Inject multiple kinds of knowledge by training adapters independently on different tasks, support continual knowledge infusion

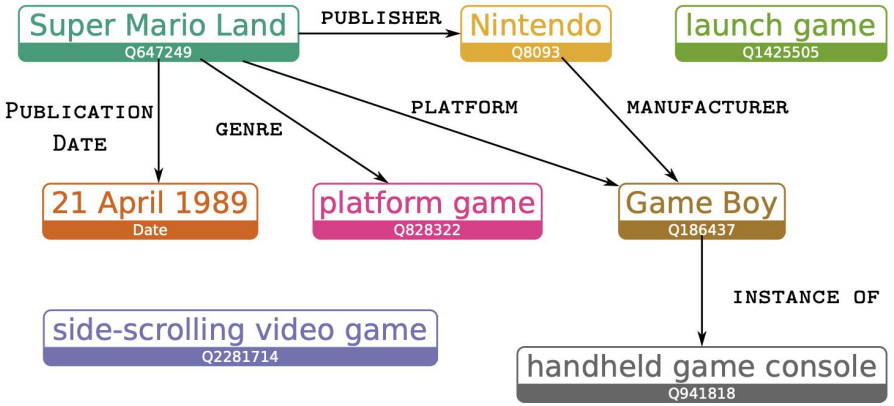


Ruize Wang, Duyu Tang, Nan Duan, Zhongyu Wei, Xuanjing Huang, Jianshu ji, Guihong Cao, Daxin Jiang, Ming Zhou. K-Adapter: Infusing Knowledge into Pre-Trained Models with Adapters. Arxiv:

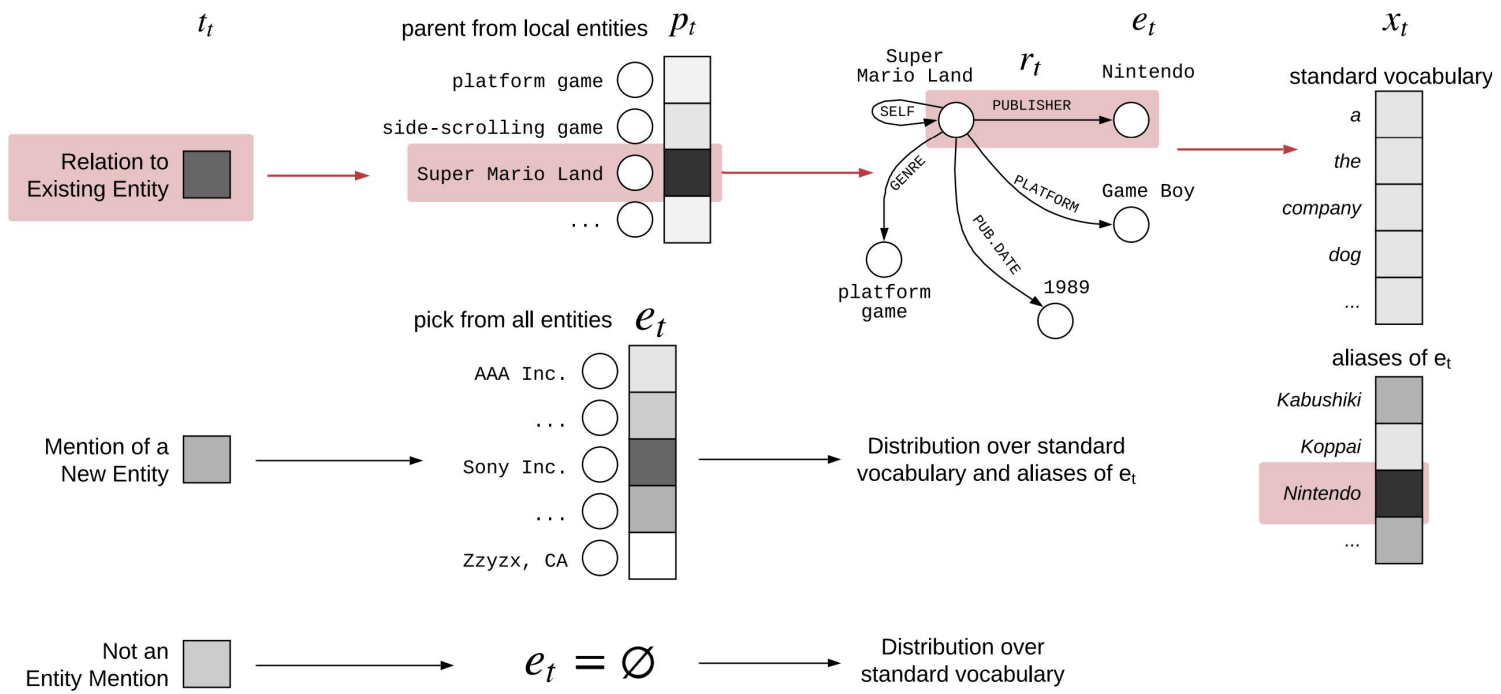
Knowledgeable Framework

- LM with mechanisms for selecting and copying facts from KG

[*Super Mario Land*] is a [*1989*] [*side-scrolling*] [*platform video game*] developed and published by [*Nintendo*] as a [*launch title*] for their [*Game Boy*] [*handheld game console*].



Super Mario Land is a 1989 side-scrolling platform video game developed and published by Nintendo



Robert L. Logan IV, Nelson F. Liu, Matthew E. Peters, Matt Gardner, Sameer Singh. Barack's Wife Hillary: Using Knowledge-Graphs for Fact-Aware Language Modeling. ACL 2019.

Resource: Chinese Pre-Trained Models（CPM）

训练数据	模型大小			任务
新闻				文本分类
百科	参数量			自然语言推理
对话	109M	334M	2.6B	阅读理解
网页	层数			完形填空
故事	12	24	32	对话生成
	隐向量维度			实体生成
	768	1,024	2,560	
	每层注意力数			
	12	16	32	
	注意力向量维度			
	64	64	80	

```
print("输出:",tokenizer.decode(generates))
```

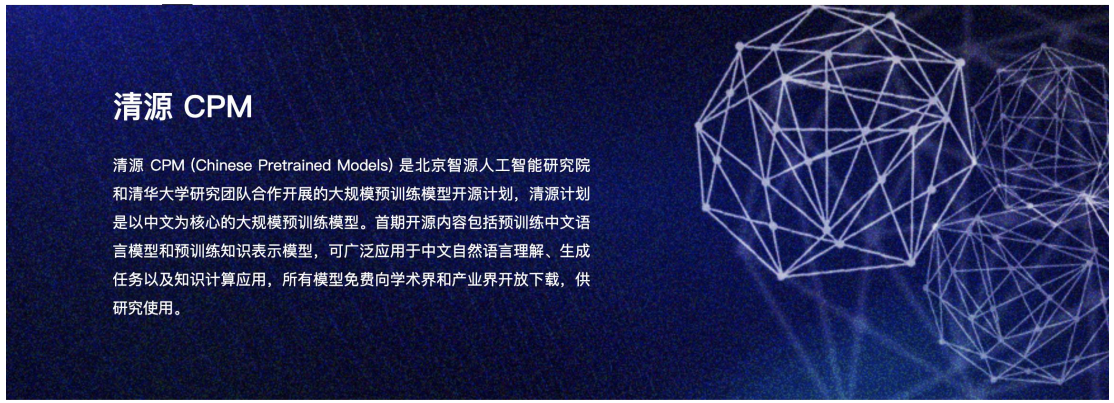
输入：我们当中要数小明要厉害->小明
王文昨天去了上海，之后又回到了武汉->王文
他看到一个人，那人正是他的老师周治平->周治平
我们都认为欧阳一凡的为人不错->

输出：欧阳一凡

输入：
姚明的身高是211cm，是很多人心目中的偶像。->姚明,身高,211cm
毛泽东是绍兴人，早年在长沙读书。->毛泽东,出生地,绍兴
虽然周杰伦在欧洲办的婚礼，但是他是土生土长的中国人。->周杰伦,国籍,中国
小明出生于武汉，但是却不喜欢在武汉生成，长大后去了北京。->小明,出生地,武汉
吴亦凡是很多人的偶像，但是他却是加拿大人，另很多人失望->吴亦凡,国籍,加拿大
武耀的生日在5月8号，这一天，大家都为他庆祝了生日。->武耀,生日,5月8号
《青花瓷》是周杰伦最得意的一首歌。->周杰伦,作品,《青花瓷》
北京是中国的首都。->中国,首都,北京
蒋碧的家乡在盘龙城，毕业后去了深圳工作。->蒋碧,籍贯,盘龙城
上周我们和王立一起去了他的家乡云南玩，昨天才回到了武汉。->王立,籍贯,云南
昨天11月17号，我和朋友一起去了海底捞，期间服务员为我的朋友刘章庆祝了生日。->

输出：刘章,生日,11月17号

Resource: Chinese Pre-Trained Models（CPM）



CPM-Generate

Chinese Pre-Trained Language Models (CPM-LM) Version-I

Python MIT 54 595 9 0 Updated 2 days ago

arXiv:2012.00413 [pdf, other] cs.CL

CPM: A Large-scale Generative Chinese Pre-trained Language Model

Authors: Zhengyan Zhang, Xu Han, Hao Zhou, Pei Ke, Yuxian Gu, Deming Ye, Yujia Qin, Yusheng Su, Haozhe Ji, Jian Guan, Fanchao Qi, Xiaozhi Wang, Yanan Zheng, Guoyang Zeng, Huanqi Cao, Shengqi Chen, Daixuan Li, Zhenbo Sun, Zhiyuan Liu, Minlie Huang, Wentao Han, Jie Tang, Juanzi Li, Xiaoyan Zhu, Maosong Sun

Abstract: ...as the training corpus of GPT-3 is primarily English, and the parameters are not publicly available. In this technical report, we release the Chinese Pre-trained Language Model (**CPM**) with generative pre-training on large-scale Chinese training data. To the best of our knowledge,... ▽ More

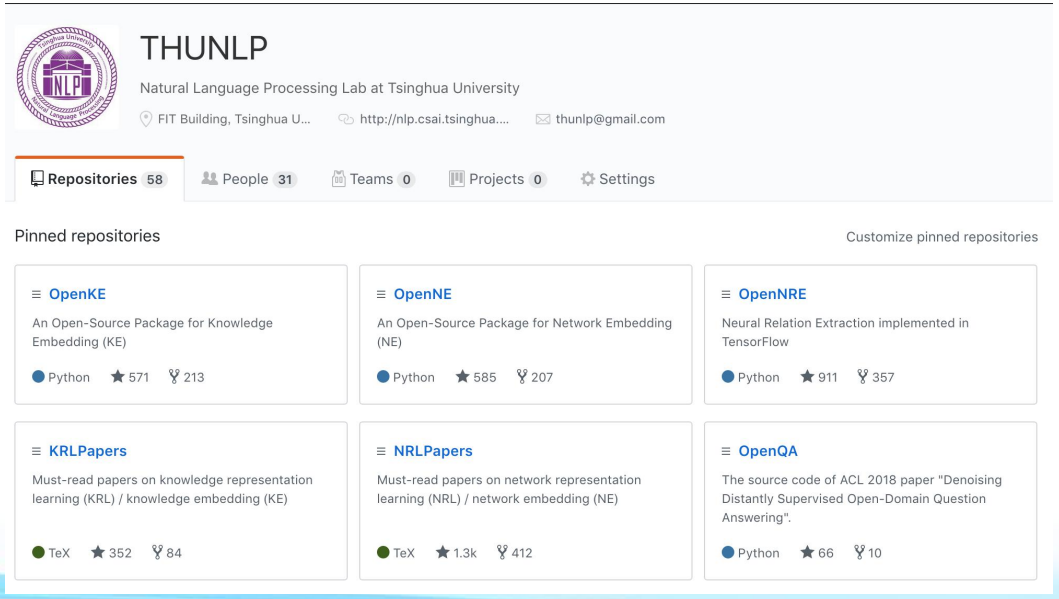
Submitted 1 December, 2020; **originally announced** December 2020.

主页（含模型下载）　源码　技术报告

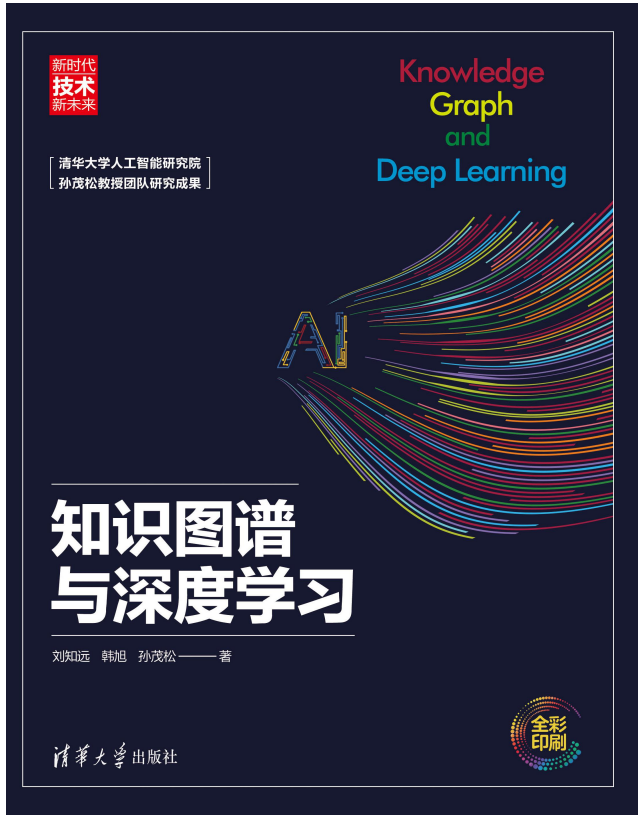
Open Source

- Packages for representation and acquisition of linguistic and world knowledge
- The projects obtain 40000+ stars on GitHub

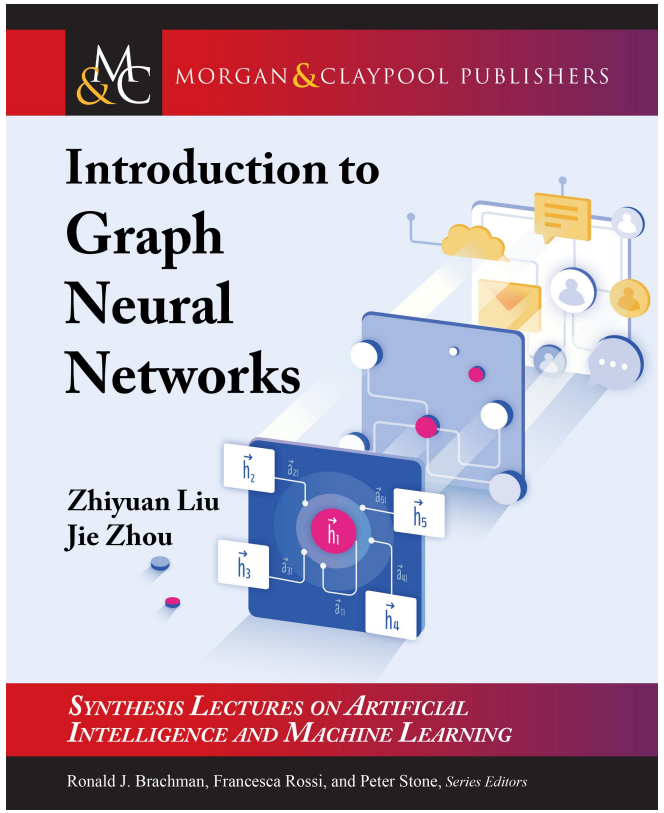
https://github.com/thunlp



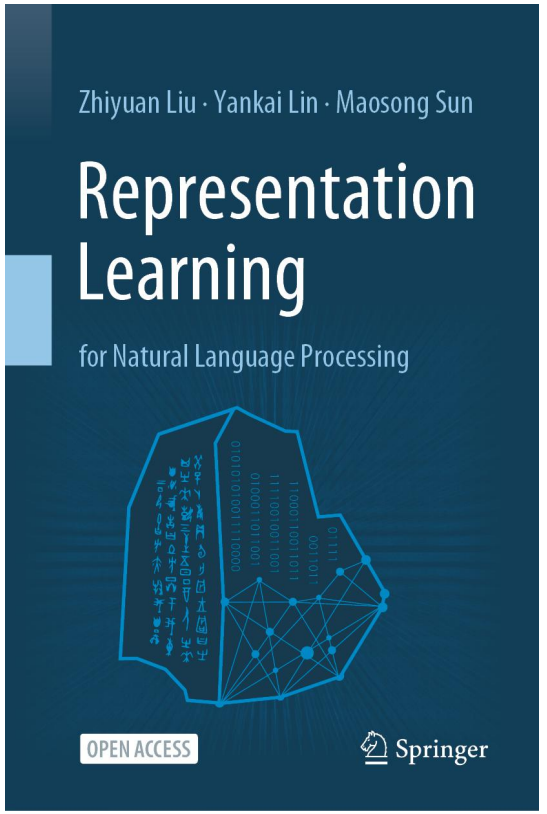
Books



KG and DL



GNN

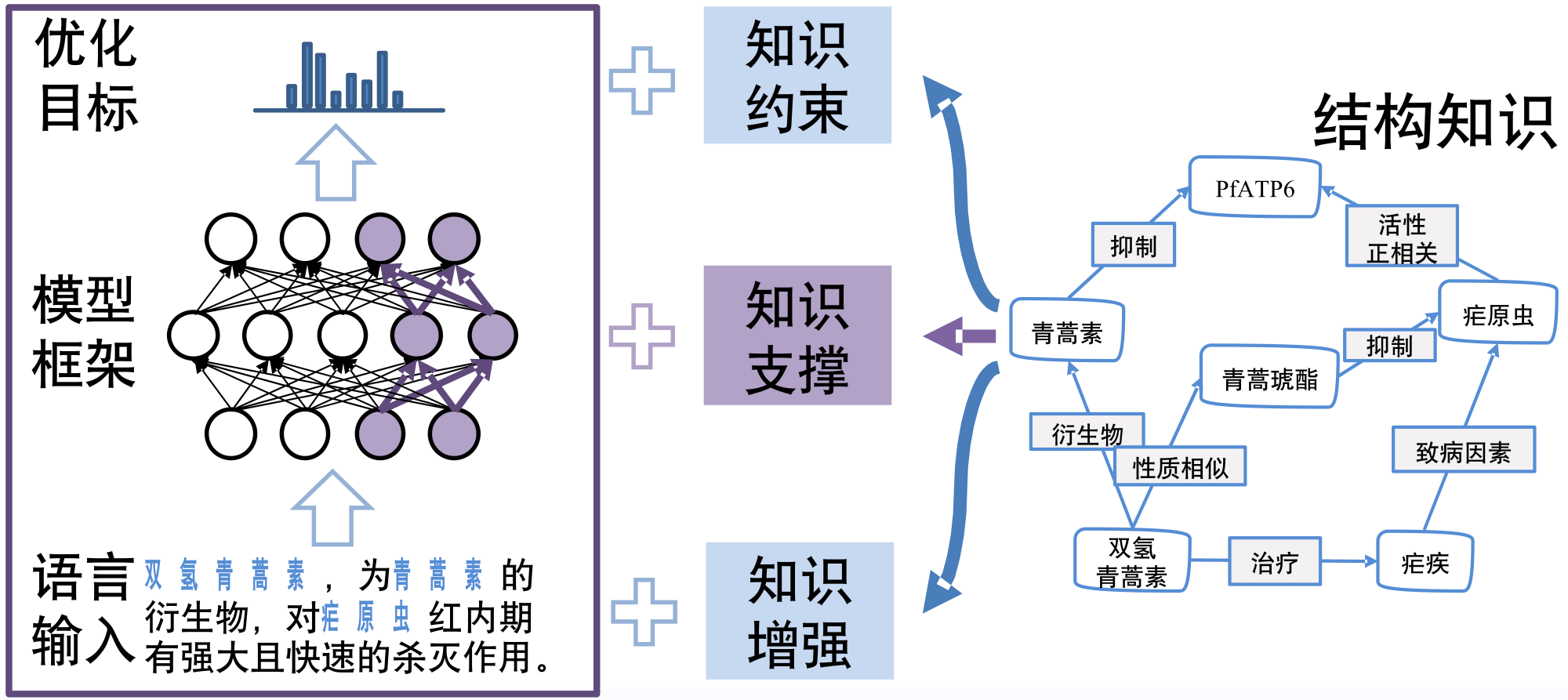


RL for NLP

Open Access!

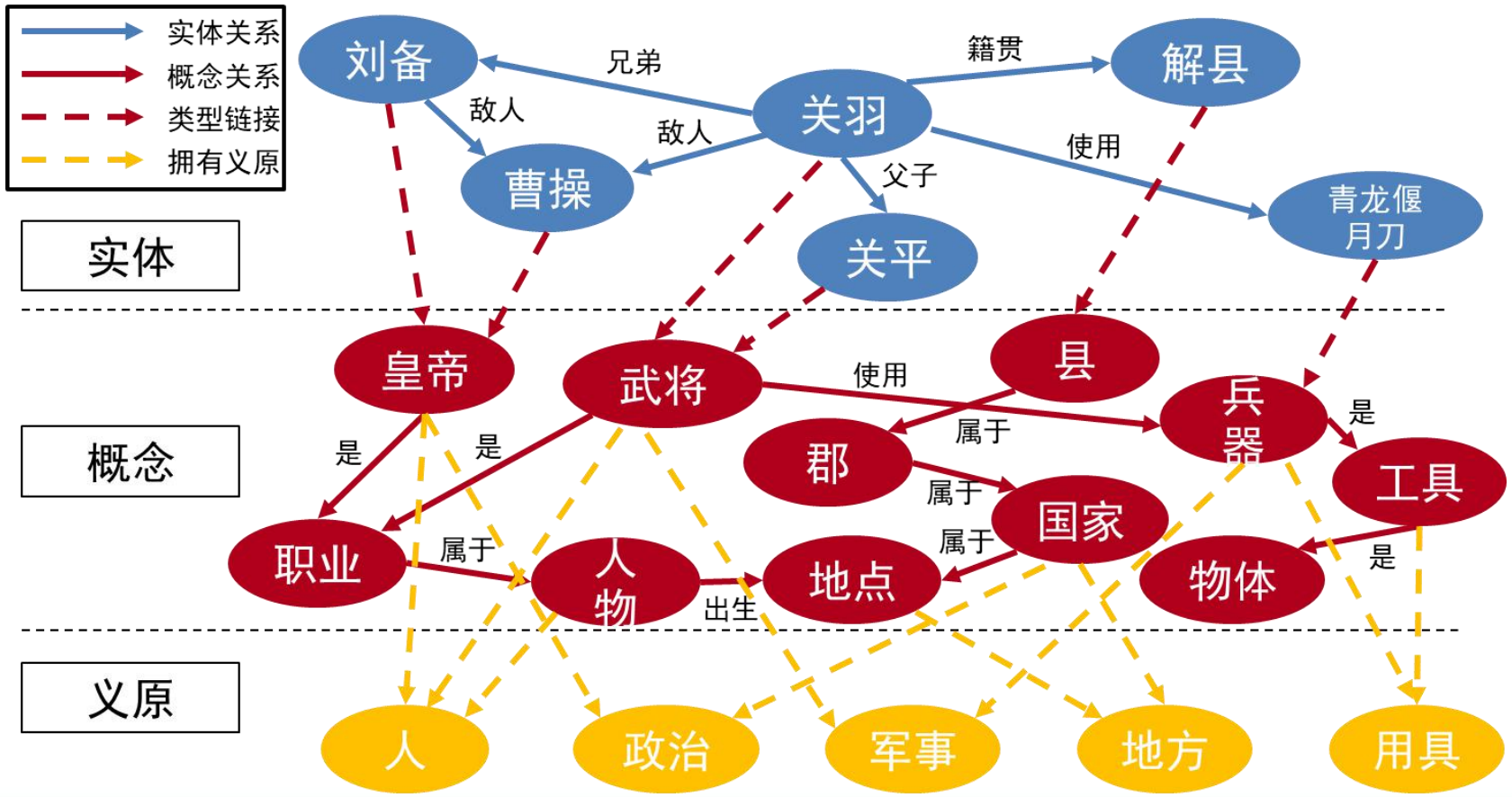
Outlook

- More methods to incorporate multiple knowledge into deep learning

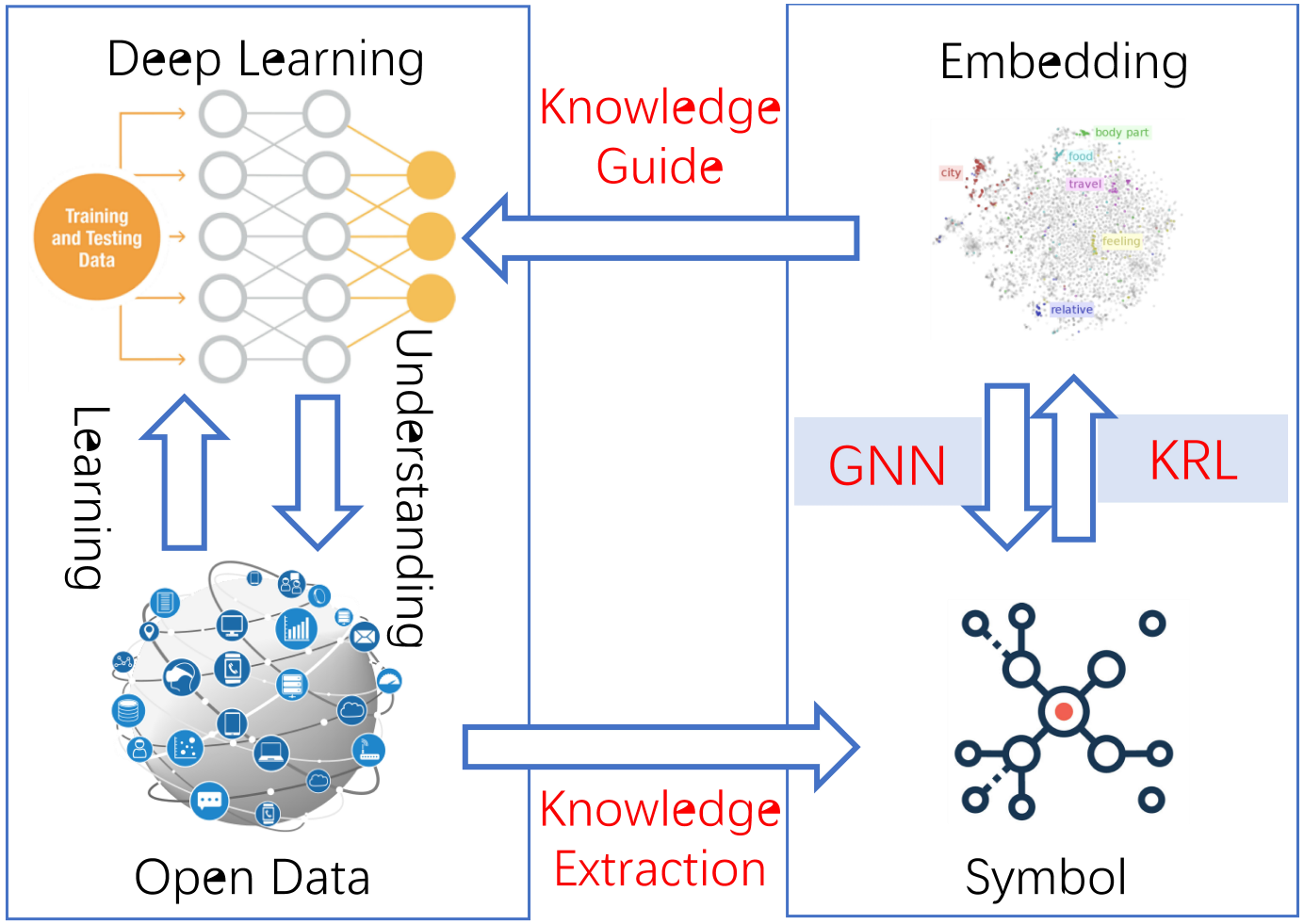


Outlook

- More knowledge in future, concepts, commonsense, event, …



Knowledge-Guided NLP



Deep Learning Knowledge Graph

Special Issue CFP on Pre-trained Language Models

http://www.keaipublishing.com/en/journals/ai-open/

Guest Editor: Zhiyuan Liu, Xipeng Qiu, Jie Tang

AI Open Special Issue/Section on Pretrained Language Models
Call for Papers

The release of ELMo, BERT, and GPT in 2018 indicates the success of pre-trained language models (PLMs), and the following years witness their great breakthrough on natural language understanding and generation. Many works have been done to explore more efficient and effective architectures for pre-training, to further improve pre-trained language models with cross-modal data, cross-lingual data, and structured knowledge, etc., or innovatively apply PLMs in various NLP-related tasks.

This special issue on Pretrained Language Models is devoted to gathering and presenting cutting-edge review, research, or applications of PLMs, providing a platform for researchers to share their recent observations and achievements in this active field. Specific topics for this special issue include but are not limited to:

- Novel architectures and algorithms of PLMs
- Generative PLMs
- Fine-tuning and adaptation of PLMs
- Multi-task and continual learning of PLMs
- Knowledge-guided PLMs
- Cross-lingual or multi-lingual PLMs
- Cross-modal PLMs
- Knowledge distillation and model compression of PLMs
- Analysis and probing of PLMs
- Applications of PLMs in various areas such as information retrieval, social computation, and recommendation

Thanks!

liuzy@tsinghua.edu.cn

http://nlp.csai.tsinghua.edu.cn/~lzy